I’ve been asked to address the question of whether the federal government can “ensure that all states adopt rigorous, broad educational standards and aligned curricula and assessment instruments.” If the answer is “no” – and the evidence to date certainly does not encourage optimism on this question – I am then asked to comment on the implications that follow.

Let me begin by addressing a prior question: why does it matter if states adopt rigorous standards and well-aligned assessments? To address this question, I need to go back to first principles and articulate the core assumptions underlying the standards movement. Although I would like to believe these are by now widely understood and shared, it is easy to lose sight of them in the arguments that rage over testing, accountability and NCLB.

In its simplest terms, the standards movement attempts to apply well-established principles of effective organizational development and behavior to the K-12 sector. Effective organizations have focused missions; articulate clear, measurable short and long-term goals; align their financial and human resources to accomplish their goals; measure progress regularly against their goals and make mid-course corrections where necessary based on data; empower line managers with the resources and authority to direct the core work of the organization; and hold managers accountable for results.

The architects of the standards movement argued that K-12 policymakers first needed to come to agreement that the core mission is an academic one: namely, to prepare all students with the foundation of knowledge and skills required for further learning, work, and citizenship. Given the rapidly shrinking proportion of family-wage jobs available to people with only a high school education, the “further learning” part of
the mission statement has come to mean preparing all students to be “college-ready” – i.e. to be able to go on to post-secondary education if they choose to do so.

In 1990 the nation’s governors and President G.H.W. Bush adopted a first-ever set of National Education Goals, the most important of which stated that “by the year 2000, all students will leave grades 4, 8, and 12 having demonstrated competency over challenging subject matter, including English, mathematics, science, foreign languages, civics and government, economics, arts, history and geography … ,” Leaving aside for a minute the grandiosity of these ambitions, this goal became in effect a new academically-focused mission statement for public education, and the early standards development work already undertaken by a handful of states and by the National Council of Teachers of Mathematics suggested a way to operationalize the mission statement. By the end of the 1990’s virtually all states had developed learning standards in the four core academic subjects for each level of schooling, and most had adopted assessment systems and accountability policies designed to ensure that mastery of the state’s learning standards became the core goal of the enterprise.

The adoption of learning standards and assessments by the states carried very substantial equity implications. In a pre-standards world in which individual districts were free to set their own learning goals for students, and districts and schools were free to establish different learning goals for different groups of students, race and class were the primary predictors of the quality and rigor of the academic program offered to students, and English language learners and students with disabilities were almost universally exposed to less challenging academic content. The persistence today of widespread achievement gaps between white and non-white students, and between “regular” education students and ELL and special needs students, only confirms the obvious point: the adoption of common, rigorous learning standards for all students is but a starting point, necessary but by no means sufficient to meet the goal of enabling all students to achieve at high levels. But the creation of common learning standards that all students are expected to meet by the end of high school offers the most powerful lever yet available to equity advocates to assert the state’s obligation to provide adequate
learning opportunities for all students, as the Campaign for Fiscal Equity has demonstrated in its successful litigation against the State of New York. It is for this reason that some standards advocates describe the movement as having created a new “civil right” – the right to be taught to common high standards -- for those who have historically been least well served by our schools.

The Quest for Rigorous Standards and Aligned Tests

Virtually from the moment the governors and President Bush I adopted the National Education Goals, the question of how best to pursue national goals in the context of a highly decentralized, state-based education system has been on the table. Within a year after the adoption of national goals, then-Secretary Lamar Alexander worked with Congressional leaders to establish the National Council on Education Standards and Testing (NCEST), a bi-partisan group charged with reporting back to Congress on the desirability and feasibility of establishing national standards and assessments. NCEST members struggled to find the appropriate balance between the need for a more national framework within which the country could pursue the attainment of the very ambitious national goals that had just been set, and the need to respect the sovereignty of states in setting educational policy. In its 1992 report, NCEST endorsed the need for national standards, emphasizing that these should be voluntary and national (as distinct from federal) but backed away from the Bush Administration’s proposal for a single national test, calling instead for a system of national tests from which states might choose depending on the nature of their curriculum. This convoluted, Rube Goldberg-like solution to the challenge of providing national guidance without federal control foreshadowed the political difficulties that befell subsequent efforts over the next 15 years to impose some form of national quality control over the standards and assessment development work of states.

It is important to remind ourselves that the goal of ensuring that “all states adopt rigorous, broad educational standards and aligned curricula and assessment instruments” has been federal policy since 1994, when a provision to this effect was written into that
year’s reauthorization of the Elementary and Secondary Education Act (called Improving America’s Schools Act). States were given a date certain by which such standards and aligned assessments were to be in place: failure to comply meant jeopardizing federal aid. Similar language was carried over into the No Child Left Behind Act (2002). To the best of my knowledge no state has yet been penalized for being out of compliance with this provision, and yet by the testimony of the three national organizations that have been reviewing and analyzing state standards and tests over this period, there continues to be enormous variation in the quality and rigor of state standards, and in the degree to which state assessments measure the full range and depth of those standards.

The American Federation of Teachers (AFT) and the Thomas B. Fordham Foundation have been reviewing the quality of state standards since the late Nineties, with similar stated criteria. Both favor standards that are grounded in the core disciplines, focus on essential knowledge, are clear and specific, and make difficult choices about what to leave out. Both organizations have recently revisited the quality of state standards for the first time in several years. In 2000 Fordham’s average state grade across all subjects was C-. Six years later their average grade remains the same, despite the fact that 37 states have made revisions in at least one set of standards during this period. While Fordham reviewers credit some states, including New York, with having made significant improvements during this period, others in Fordham’s view have gone backwards. Fordham’s bottom line is that only three states deserve an A for the quality of their standards – California, Indiana, and Massachusetts; six more (including New York) deserve some form of B; the rest, more than half, get D’s and F’s.

It should be no surprise that the AFT’s grading scale is a bit more lenient, and that the teachers’ organization gives the states more credit for progress than does Fordham. In its 2006 report, the AFT focused on alignment, looking for states that in its judgment had both strong content standards and tests that were well aligned with those standards. The AFT credits 11 states with meeting these criteria. Four of those states – California, Indiana, New York, and Virginia -- are Fordham A or B states; but two of AFT’s other stars – Ohio and Washington – rate no better than D by Fordham’s criteria. At the
bottom of both lists are states like Arkansas, District of Columbia, Hawaii, Montana, North Dakota, and Wyoming, along with one very surprising anomaly: Connecticut, a state that normally shows up on lists of high-performers. This suggests an important caution about the degree of weight to place on quality of standards in evaluating the overall quality of a state’s improvement strategy, a point I will return to below.

The third organization that has issued reports on the quality of state standards and tests, Achieve, was established in 1997 by a bipartisan group of governors and corporate CEOs principally to provide advice and assistance to states as they were developing their standards, assessments, and accountability systems. In its first several years Achieve’s core activity was its Benchmarking Initiative, a process designed to provide in-depth analyses of standards and tests for interested states. Achieve did not rank or rate states, but rather provided them with the results of a very close analysis of the quality of their standards benchmarked against the very strongest US and international standards, followed by a careful review of the state tests. While crediting most states with moving away from a reliance solely on multiple choice questions and adopting a better mix of test formats, Achieve found that in general most state assessments did a much better job of measuring student performance on the low end of state standard (i.e. basic skills and knowledge) than assessing the more ambitious kind of reasoning, problem-solving, and analytic skills that most state standards documents call for.

While it may not be realistic to expect a federal review process to be able to undertake the kind of in-depth alignment analysis that Achieve was able to conduct at the request of states, it is important to at least understand the big questions that a serious alignment analysis is designed to answer. Achieve’s alignment review began with a close examination of the test blueprint, attempting to map each test question against at least one standard. Then Achieve’s reviewers examined each item along three dimensions: “content centrality” (i.e. does the content of the question match the content of the standard?); “performance centrality” (i.e. does the type of performance presented by the item match the type of performance described in the corresponding standard?); “challenge” (i.e. does the level of challenge of the set of items related to a particular
standard represent an appropriate range of difficulty for the particular grade level?). Finally, the reviewers looked at “balance and range,” asking whether the test taken as a whole gauges the depth and breadth of the standards. The bottom line from Achieve’s Benchmarking Initiative, at least at the time I left the organization at the end of 2002, was that of the 14 states for whom we had undertaken such an analysis, only one – Massachusetts – in our judgment had rigorous, high quality standards and high quality tests that were carefully aligned with its standards. Achieve’s benchmarking work has shifted focus in the last four years, as I will explain below, but it is important to point out that several states, most notably Indiana, acted on the results of Achieve’s early analyses to significantly strengthen their standards and tests, which was the point of the Benchmarking Initiative.

Before returning to the core question of the feasibility of expecting the federal government to “ensure” rigorous, well-aligned state standards and tests, I want to stand back a bit from the review processes and findings of Achieve, AFT, and Fordham and make some more general observations about the quality of standards and tests. In 2002 Marshall Smith and I wrote a piece entitled, “Staying the Course with Standards-Based Reform: What It Will Take.” This was part of an edited volume of reflections from several members of the Pew Forum on Standards-Based Reform, a floating seminar that brought together leading policymakers, researchers, and practitioners to analyze (and to some degree, help shape) the standards movement as it was developing during the Nineties. Smith and I outlined three sets of questions that we thought state policymakers should be asking about their standards, based on our own observations in schools as well as the analyses of the three reviewing organizations. These questions seem as pertinent to me today as they did four years ago.

- Do the knowledge and skills we are asking students to master reflect the most essential concepts in this field – what all students really need to know and be able to do – or do they reflect the collective wish lists of subject matter specialists?
• Are these standards grounded in the real-world requirements for effective citizenship, employability in a changing economy, and post-secondary education?

• Are these standards, taken as package, manageable in the classroom? Could, for example, a conscientious, well-prepared elementary school faculty, working within the constraints of the normal school calendar, be expected to enable virtually all of its students, however diverse their abilities, to master the material embodied in the standards by the end of grade 5? If so, would there still be room for the teachers to include some topics of their own choosing that might fall outside these standards?

On tests, Smith and I warned against the continuing practice in some states of using off-the-shelf tests to measure student progress on state standards, but we also encouraged states to work with districts to help them develop more instructionally useful assessments to supplement the state tests. In this regard there has been substantial progress in the last four years, as districts have begun to invest in the development of formative assessments and teachers have begun to see the benefits of having access to more timely and instructionally sensitive information about skills and concepts students are struggling with as well as those they are mastering.

**Previous Federal Attempts at Quality Control**

It is important to remind ourselves that at the outset of the standards movement the hope was that we could forge enough of a political consensus around national standards so that, even if they were voluntary, virtually all states would use them as guideposts in the development of their own standards. In fact this is what happened in mathematics where, at least until the outbreak of the math wars in California, virtually all states adopted nearly in toto the standards developed by the National Council of Teachers of Mathematics. Unfortunately, the well-intentioned efforts on the part of the Bush I administration to encourage the development of similar standards in English and history backfired, leading to an environment for the next decade in which it was nearly
impossible for elected officials to utter the words “national” and “standards” in the same sentence. Consequently, the Clinton and Bush II administrations had to resort to much more indirect strategies to address the challenge of bringing some measure of national quality control to the standards and assessment development work of states.

The first attempt to address the quality and comparability of state standards was a now-forgotten provision in the Goals 2000 Act (1994) that called for the creation of a national council to which states could voluntarily bring their standards and tests for review and comment. The National Education Standards and Improvement Council (NESIC), whose members were to be appointed by the President and Congress, was designed to help ensure that the standards and tests being developed by the states with Goals 2000 funding were of high quality. Governors like Roy Romer (CO) and John Engler (MI) spoke at the time of how useful it would be for them to have access to an external review process in order to verify the claims being made to them by their education leaders that the standards under development in their states were in fact “world class.”

Unfortunately, the proposed Council was an early casualty of the Republican takeover of Congress in 1995. Conservative critics attacked NESIC as an attempt to create a national school board. Consequently, the Clinton administration never bothered to submit its proposed slate of nominees to the Council, and this provision of Goals 2000 was quietly repealed. The desire of governors like Romer and Engler for such a review service did not go away, however, and a year later they and other governors and corporate leaders decided to create an independent, non-profit organization to carry out this function, which explains how and why Achieve came into existence.

The next federal attempt to address the quality and comparability of state standards came at the beginning of Clinton’s second term, when the President proposed the development of voluntary national tests in reading in grade 4 and math in grade 8. The unstated premise behind this proposal was that, even though these exams were to be voluntary and have no stakes attached, parents would be motivated to take action if the
information they received about their child’s achievement from the national test was wildly at variance with what they were told from the state test. Under the most optimistic scenario the voluntary national tests would acquire sufficient political salience to motivate states to use them as benchmarks against which to align their own standards and tests. Again, however, a Republican-controlled Congress opposed the Clinton proposal as an undue expansion of the federal role, and the program died for lack of Congressional authorization.

The most recent attempt to address the quality control problem is the provision in NCLB requiring states to participate in the National Assessment of Educational Progress (NAEP). The premise underlying this strategy is similar to that behind the proposed voluntary national test: namely, that undue discrepancies between performance on state tests and NAEP would motivate parents to take action to raise standards. Because NAEP is administered on a sampling basis and does not yield individual student scores, it is hard to see how this strategy will have the desired political effect. Thanks to the work of Mark Musick at the Southern Regional Education Board, the disparities in many states between the proportion of students deemed proficient on NAEP tests versus those deemed proficient on state tests have been known for many years, to little political effect, so it seems unlikely that, absent information to parents about their own child’s or school’s performance on NAEP, NAEP results will serve as a goad to state action to raise standards. In summary, then, the evidence to date is that eleven years after federal law first required states to adopt rigorous standards and aligned assessments, we are still very far from that goal.
Rethinking NCLB

For all its problems, NCLB has accomplished one extremely important objective: it has forced states to redefine a successful school from one in which on average students are learning well, to one in which all groups of students are learning well. This was an intended consequence of the law, and whatever else may change in the next reauthorization, the disaggregated reporting requirements of NCLB must be retained.

An important unintended consequence of NCLB has been to bring back on the table for public discussion the case for national standards. The “adequate yearly progress” provisions of the law have brought to the surface the absurdity of attempting to impose a federal accountability template on 50 states after each state has developed its own standards, adopted its own assessments, and set its own definition of proficiency. This has created incentives for states to game the system by lowering the bar, and it has led to a situation in which states with weak standards and low levels of student performance are deemed successful (i.e. they have few failing schools or districts), while states with more rigorous standards and higher levels of student achievement have much larger proportions of schools deemed failing under NCLB.

Before taking up the issue of national standards, let me make a modest proposal for the next round of NCLB discussions. The current NAEP requirements may not be working as the drafters hoped to shame states with low standards into strengthening them, but what if performance on NAEP were to be the new accountability measure for states. If states were required to show progress every two years on NAEP, it would likely have the effect of motivating states to bring their own state assessment systems more into alignment with NAEP. I realize that this is not how NAEP was intended to be used, but NAEP tests have broad legitimacy among education policy leaders, and at least a handful of states have already seen the virtue of aligning their state tests with NAEP.

In addition to the general concern about using NAEP for accountability purposes, I can anticipate at least two additional concerns that critics might raise. The first has to
do with NAEP’s performance levels, more specifically its overly ambitious definition of proficiency. This concern is at the heart of Richard Rothstein’s conference paper. My proposal would require the establishment of bi-annual improvement targets on NAEP that would be derived from each state’s baseline performance and, following Bob Linn’s suggestion, from a careful analysis of the actual improvement rates of benchmark states, not on some artificial target rate that no state or nation has ever achieved. With Rothstein, I would abandon the fiction of a one hundred percent proficiency goal, especially if we are using the NAEP measure of proficiency.

The second concern has to do with the design of NAEP tests, which are based on a sampling technique which provides information at the state level but does not produce scores for individual districts, schools, or students. If NAEP is now to be used to hold states accountable for performance, how will states persuade districts and school to motivate students to take the test seriously when the results don’t really count for them? This may not be a problem for elementary or middle school students, but I think there is good reason to be nervous about attaching state accountability to a test that high school students have no reason to take seriously. I think we need a separate strategy for holding states accountable for high school improvement, as I will outline below in my discussion of the American Diploma Project.

Just to be clear, I am proposing that federal AYP requirements ought to apply only to states, not to districts or schools. NCLB should continue to require states to have their own accountability systems and adequate progress requirements, but we should end the public confusion of publishing separate (and often conflicting) state and federal ratings of schools. For better or worse we have a state-based education system. The challenge in our system for each level of government is to craft a set of policies that combine the right mix of guidance, pressure and support to induce the level just below it to do the right thing. The federal focus ought to be on helping states strengthen their capacity to provide appropriate guidance, pressure, and support to districts, and to monitor the performance of states to ensure that all groups of students are making progress.
**National, Not Federal Strategies**

Beyond NCLB, what might we do to create a more national set of strategies to ensure all students the right to be taught to comparably high standards? The Fordham Foundation recently released a report outlining four possible approaches to the problem. One approach would have the federal government create national standards and tests. In my view this would be bad policy as well as bad politics. Another approach would involve “sunshine and shame,” i.e. using NAEP to embarrass states with low standards into raising them. As I argued earlier, I see no evidence that this approach will work. Their other two suggested approaches, voluntary national standards and voluntary consortia of states, strike me as having more promise, so let me outline some ways in which these strategies might work.

One way to address the problem of national standards is to encourage, with or without federal support, national organizations like Fordham, AFT, and Achieve to come forward with proposed sets of model national standards. These organizations have well-developed views about what good standards look like, and states would likely pay attention to their proposals. If performance on NAEP were to become the accountability measure for state progress, such standards would need to be aligned with NAEP in order to be taken seriously by states.

A second, related strategy is to encourage organizations with curriculum and assessment development capacity to come forward with a broad instructional guidance package that would include standards, aligned curriculum, and formative and summative assessments. The College Board is one such organization that is already in this business; the National Center on Education and the Economy, which in partnership with the University of Pittsburgh’s Learning Research and Development Center produced the best set of national standards in the 1990’s, is another such candidate. The argument for encouraging organizations like these two, or the three mentioned just above, to take on the task of standards development is that they are independent, non-governmental
organizations with a clear point of view. Consequently, they are much likelier than states or the national subject matter organizations to come up with standards that are coherent across content areas, focus on essentials, and are manageable in the classroom.

The most promising example of the voluntary state consortium approach is the American Diploma Project (ADP). ADP was launched in 2002 by Achieve, The Education Trust, and the Thomas B. Fordham Foundation. The initial goal of the project was to help states align their high school exit standards and graduation requirements with the expectations of higher education institutions and high-performance employers. Although one might have reasonably inferred that the standards development process in most states would have begun by asking, “What must all students know and be able to do upon leaving high school?” and then mapped backward to assure that the standards at key grade levels would ensure that students were on track to meet those end goals, there is very little evidence that most states proceeded in this fashion. Consequently, ADP began by working with higher education officials and K-12 leaders in five states to analyze the alignment between high school graduation requirements and college admissions and placement standards in those states, as reflected in the assessment instruments used for these purposes by each system. Simultaneously, ADP launched a national study of a cross-section of high-performance worksites to ascertain the kinds of reading, writing, and math skills employers expected their entry-level hires to possess. Working from these two sets of data, ADP staff developed a set of benchmark standards against which states might measure the rigor and appropriateness of their own high school standards and exit requirements.

The work of ADP has helped raise consciousness among state policymakers about the “expectations gap” between high school exit standards and college and workplace readiness requirements. In the wake of the high-profile 2005 National Summit on High Schools co-sponsored by Achieve and keynoted by Bill Gates, 26 states have now joined with Achieve to form the ADP Network. These states have agreed to work together to take advantage of the ADP benchmark standards to align their high school standards,
assessments, graduation requirements and accountability systems with the demands of higher education and the workplace.

About a year ago Achieve conducted a baseline survey of all 50 states on the four commitments required to join the ADP Network. On the first, aligning high school standards with the ADP benchmark standards, only five states reported having completed the process, but another 30 reported movement in that direction. On the use of state test results for college admission or placement purposes, six states reported that they are currently doing this, with eight more planning to do so. On the alignment of high school graduation requirements with the expectations of colleges and employers, eight states met Achieve’s criteria, which require all students to complete four years of rigorous English courses and a sequence of math courses that includes Algebra II; another 12 states plan to adopt such graduation requirements in the future. Finally, on the fourth commitment, holding high schools accountable for improving the college and work readiness of their students, four states said they have such policies in place, and nine more plan to do so. An important prerequisite of this last commitment is the creation of a unified P-16 longitudinal data system which allows the tracking of students into post-secondary education. Only three states currently have such a system in place, but 31 reported that they were in the process of either linking their two data systems or creating a unified new system.

The ADP Network is potentially important for three reasons. First, it focuses the attention of state policymakers on the right goal: graduating all kids college-ready. Second, it encourages states to think in a more unified way about aligning their secondary and post-secondary systems, and to use student performance data from the post-secondary system to continuously improve the quality and rigor of their high school programs. Third, it is already prompting some states to consider investing together in the development of common end-of-course high school tests to help ensure that all students taking courses with the same label are in fact getting comparably challenging content. Nine states have already agreed to jointly sponsor the development of an Algebra II test that all can use. If this initiative is successful, other states and other subjects will follow.
The ADP Network points us toward an alternative set of measures for holding states accountable for the performance of high schools, given the problem cited earlier with NAEP. If the target for high schools should be all students graduating college-ready, then the relevant measures for gauging bi-annual improvement should include such factors as the proportion of students completing a college-ready curriculum, pass rates on state end-of-course or other college-aligned tests, high school graduation rates, and post-secondary enrollment and success rates. It may seem utopian to imagine federal accountability requirements for states that would bridge into their post-secondary systems, but given the recent report of the Spellings Commission, the time may be right to at least introduce the idea of an aligned K-16 approach to accountability for student success.

Concluding Thoughts

The big achievement of the standards movement over the past 15 years has been to remind us that the core mission of our schools is an academic one, and to redefine the core goal as one of enabling all students to leave high school with a solid foundation of college and work-ready knowledge and skills. The main achievement to date of NCLB has been to cast a searchlight on the performance of key subgroups of students and to compel educators to pay special attention to those groups of students who have historically been least well served by our schools. If one is in any doubt about the equity consequences of these shifts in focus and priorities of public policy toward the schools, think about the performance and public standing in the early Nineties of such major urban districts as New York, Chicago, Philadelphia, San Francisco, and Boston, compared to the performance and public perception of these districts today. These districts now have rates of improvement in student performance that typically exceed their state averages, and they are seen increasingly by education leaders as beacons of innovation rather than bureaucratic quagmires. Large-scale improvement in districts like these stems from many factors, not the least of which are changed governance arrangements and stable, highly skilled leadership. But the standards movement has created an accountability environment that has encouraged policymakers to shift attention and resources to the
schools and students serving the neediest students, and has provided incentives and leverage for district leaders to pursue an aggressive reform agenda.

The challenge for the next 15 years is to recalibrate the balance between top-down and bottom-up reform strategies, to strengthen the role of professional judgment and reduce our reliance on prescription, and to think more about carrots and less about sticks. The steps I have proposed -- making state accountability the focus on NCLB by using NAEP as the measure of adequate progress; encouraging independent national organizations to come forward with proposed sets of national standards aligned with NAEP; supporting the development of consortia of states that are committed to the goal of college readiness for all -- might free up space for state and local leaders to pay more attention to capacity-building, and especially to strategies to support the professionalization of teaching. This brings me back to Connecticut, the anomalous state with weak standards but good performance. Connecticut is one of two states in the country that has placed teacher policy at the center of its reform strategy, and has paid sustained attention to attracting, developing, and retaining a quality teaching force. Obviously, I’m not encouraging states to choose investing in teachers over investing in high quality standards: one needs both. But the Connecticut story is an important reminder that good teachers are the heart of the enterprise, and that high quality standards and aligned tests are only one piece of the puzzle.